

WEST Search History

DATE: Monday, August 26, 2002

| <u>Set Name</u> | <u>Query</u> | <u>Hit Count</u> | <u>Set Name</u> |
|--|--|------------------|-----------------|
| side by side | | result set | |
| <i>DB=USPT,PGPB,JPAB,EPAB,DWPI,TDBD; PLUR=YES; OP=OR</i> | | | |
| L12 | L11 and sequence.clm. and (score.clm. or confidence.clm. or probability.clm.) and (fixed.clm. or set.clm. or predetermined.clm.) | 8 | L12 |
| L11 | l7 and alignment | 377 | L11 |
| L10 | l7 and sequence.clm. and (score.clm. or confidence.clm. or probability.clm.) and (fixed.clm. or set.clm. or predetermined.clm.) | 17 | L10 |
| L9 | L8 and (fixed adj length) | 99 | L9 |
| L8 | L7 and (confidence or quality or score) | 546 | L8 |
| L7 | L6 and (protein or DNA) | 1108 | L7 |
| L6 | sequence and ((fixed or set or predetermined) adj (length or size or number)) and (representation or vector or simulation) | 29169 | L6 |
| L5 | L4 and (fixed adj length) | 9 | L5 |
| L4 | l2 and vector | 67 | L4 |
| L3 | L2 and vector and biological | 16 | L3 |
| L2 | L1 and sequence | 328 | L2 |
| L1 | kasif.in. or logan.in. or moreno.in. or suzek.in. | 2954 | L1 |

END OF SEARCH HISTORY

?ds

| Set | Items | Description |
|-----|----------|--|
| S1 | 10852110 | (SEQUENCE OR GENE OR PROTEIN OR MOTIF OR TEMPLATE) |
| S2 | 174 | S1 AND (FEATURE(W)VECTOR) |
| S3 | 19 | S2 AND ((HIGH(W)DIMENSIONAL) OR MATRIX) |
| S4 | 0 | S3 AND (SIMILAR OR HOMOLOGY) |
| S5 | 19 | S2 AND ((HIGH(W)DIMENSION?) OR MATRIX) |
| S6 | 7 | S2 AND (FIXED(W)LENGTH) |
| S7 | 2 | RD (unique items) |
| S8 | 2 | RD S6 (unique items) |
| S9 | 13 | RD S5 (unique items) |
| S10 | 918 | AU=LOGAN, B? |
| S11 | 578 | AU=MORENO, P? |
| S12 | 2 | AU=SUZEK, B? |
| S13 | 62 | AU=KASIF, S? |
| S14 | 1559 | S10 OR S11 OR S12 OR S13 |
| S15 | 0 | S2 AND ((PREDETERMINED OR SET)(W)LENGTH) |
| S16 | 30 | S14 AND (PATTERN OR EXPRESSION) |
| S17 | 27 | RD (unique items) |
| S18 | 8 | S17 AND (SEQUENCE OR GENE OR BIOPOLYMER) |
| S19 | 0 | S17 AND (COUNT? OR SCOR?) |
| S20 | 56 | AU=(LOGAN OR MORENO OR KASIK OR SUZEK) |
| S21 | 17622 | AU=((LOGAN,?) OR (MORENO,?) OR (KASIK,?) OR (SUZEK,?)) |
| S22 | 681 | S21 AND (PATTERN OR EXPRESSION) |
| S23 | 290 | S22 AND (SEQUENCE OR GENE OR BIOPOLYMER OR MOTIF) |
| S24 | 229 | RD (unique items) |
| S25 | 12 | S24 AND VECTOR |

D. AL C
(Brotech)

AN 2002116668 MEDLINE

DN 21825959 PubMed ID: 11836223

TI Classifying G-protein coupled receptors with support vector machines.

AU Karchin Rachel; Karplus Kevin; Haussler David

CS Department of Computer Science, University of California, Santa Cruz, CA
95064, USA.. rachelk@soe.ucsc.edu

SO BIOINFORMATICS, (2002 Jan) 18 (1) 147-59.

Journal code: 9808944. ISSN: 1367-4803.

CY England: United Kingdom

DT Journal; Article; (JOURNAL ARTICLE)

LA English

FS Priority Journals

EM 200206

ED Entered STN: 20020220

Last Updated on STN: 20020611

Entered Medline: 20020610

AB MOTIVATION: The enormous amount of protein sequence data uncovered by genome research has increased the demand for computer software that can automate the recognition of new proteins. We discuss the relative merits of various automated methods for recognizing G-Protein Coupled Receptors (GPCRs), a superfamily of cell membrane proteins. GPCRs are found in a wide range of organisms and are central to a cellular signalling network that regulates many basic physiological processes. They are the focus of a significant amount of current pharmaceutical research because they play a key role in many diseases. However, their tertiary structures remain largely unsolved. The methods described in this paper use only primary sequence information to make their predictions. We compare a simple nearest neighbor approach (BLAST), methods based on multiple alignments generated by a statistical profile Hidden Markov Model (HMM), and methods, including Support Vector Machines (SVMs), that transform protein sequences into fixed-length feature vectors. RESULTS: The last is the most computationally expensive method, but our experiments show that, for those interested in annotation-quality classification, the results are worth the effort. In two-fold cross-validation experiments testing recognition of GPCR subfamilies that bind a specific ligand (such as a histamine molecule), the errors per sequence at the Minimum Error Point (MEP) were 13.7% for multi-class SVMs, 17.1% for our SVMtree method of hierarchical multi-class SVM classification, 25.5% for BLAST, 30% for profile HMMs, and 49% for classification based on nearest neighbor feature vector Kernel Nearest Neighbor (kernNN). The percentage of true positives recognized before the first false positive was 65% for both SVM methods, 13% for BLAST, 5% for profile HMMs and 4% for kernNN.

09/24/2009